What is claimed is:

1. A method for extracting an attribute occurrence from template generated semi-structured document comprising multi-attribute data records comprising:

identifying a first set of attribute occurrences in the template generated semi-structured document using an ontology;

determining a boundary of each multi-attribute data record in the template generated semi-structured document;

learning a pattern for an attribute corresponding to an identified attribute occurrence of the first set in the template generated semi-structured document; and

applying the pattern within the boundary of each multi-attribute data record in the template generated semi-structured document to extract a second set of attribute occurrences.

2. The method for claim 1, further comprising the step of providing a seed ontology prior to identifying the first set of attribute occurrences.

3. The method of claim 1, wherein the ontology is one of a seed ontology and an enriched ontology.

4. The method of claim 1, further comprising enriching the ontology with the second set of attributes occurrences.

5. The method of claim 1, wherein the pattern is a path abstraction expression, wherein the path abstraction expression is a regular expression that does not comprise a union operator, and a closure operator only applies to single symbols.

6. The method of claim 1, wherein learning the pattern for each attribute occurrence comprises:

identifying the attribute occurrence in a data structure tree; and

determining the pattern of the attribute occurrence in the data structure tree.

7. The method of claim 6, further comprising the step of generalizing the pattern of the attribute occurrence prior to applying the pattern.

8. The method of claim 6, wherein the pattern comprises elements including a location and a format of the attribute occurrence.

9. The method of claim 8, wherein the elements are nodes in the data structure tree.

10. The method of claim 7, further comprising resolving the ambiguities in the extracted attribute occurrences comprising:

identifying attribute occurrences in the template generated semi-structured document matching more than one pattern;

determining a pattern that uniquely matches a given attribute occurrence and no other pattern uniquely matches the given attribute occurrence; and

eliminating matches between the given attribute occurrence and another pattern that matches the given attribute occurrence and at least one other attribute occurrence.

11. The method of claim 1, wherein learning the pattern for an attribute corresponding to an identified attribute occurrence of the first set in the template generated semi-structured document comprises:

learning positive examples of the attribute; and

learning negative examples of the attribute.

12. The method of claim 1, wherein learning the pattern for an attribute corresponding to an identified attribute occurrence of the first set in the template generated semi-structured document comprises:

determining a common supersequence for identified attribute occurrences corresponding to the attribute, wherein identified attribute occurrences are positive examples of the attribute;

determining a generalized supersequence by generalizing each term in the common supersequence; and

determining, for each term of the generalized supersequence, whether a term can be de-generalized.

13. The method of claim 1, wherein learning the pattern for an attribute corresponding to an identified attribute occurrence of the first set in the template generated semi-structured document comprises learning negative examples of the attribute, wherein the negative examples are positive examples of other attributes.

14. The method of claim 1, wherein determining the boundary of each multi-attribute data record comprises:

providing a tree of a page and a set of attribute names of a concept of the ontology;

marking a node in the tree by a set of attributes present in a subtree rooted at the node;

determining a set of maximally marked nodes in the tree;

determining a page type; and

extracting a boundary according to the page type.

15. The method of claim 14, wherein the page type is one of a home page and a referral page.

16. The method of claim 14, wherein extracting the boundary further comprises:

determining a maximally marked node with a highest score among the set of maximally marked nodes in the tree;

determining whether the tree comprises a single-valued attribute;

determining values of the single-marked attribute upon determining the single-valued attribute;

determining whether the tree comprises a multiple-valued attribute; and

determining values of the multiple-marked attribute upon determining the multiple-valued attribute.

17. A method for enriching an adaptive search engine comprising:

providing one of a seed ontology and an enriched ontology, the ontology comprising a set of concepts and a set of attributes associated with every concept;

determining an attribute identifier for a document of interest; and

adding the attribute identifier to the ontology for identifying attribute occurrences in at least the document of interest.

18. The method of claim 17, wherein determining the attribute identifier further comprises:

determining a methodology of the attribute identifier; and

determining a set of parameter values to be used by the methodology.

19. A program storage device readable by machine, tangibly embodying a program of instructions automatically executable by the machine to perform method steps for extracting an attribute occurrence from template generated semi-structured document comprising multi-attribute data records, the method steps comprising:

identifying a first set of attribute occurrences in the template generated semi-structured document using an ontology;

determining a boundary of each multi-attribute data record in the template generated semi-structured document;

learning a pattern for an attribute corresponding to an identified attribute occurrence of the first set in the template generated semi-structured document; and

applying the pattern within the boundary of each multi-attribute data record in the template generated semi-structured document to extract a second set of attribute occurrences.

20. An adaptive search engine appliance for searching a database of multi-attribute data records in a template generated semi-structured document, the search engine appliance comprising:

an ontology for identifying a first set of attribute occurrences in the template generated semi-structured document, the ontology comprising a set of concepts and a set of attributes associated with every concept;

a boundary module for determining a boundary of each multi-attribute data record in the template generated semi-structured document; and

a pattern module for learning a pattern for an attribute corresponding to an identified attribute occurrence of the first set in the template generated semi-structured document.

21. The adaptive search engine of claim 20, wherein the pattern is applied within the boundary of each multi-attribute data record in the template generated semi-structured document to extract a second set of attribute occurrences.

22. The adaptive search engine of claim 20, wherein the database of multi-attribute data records is stored on a server connected to the adaptive search engine application across a communications network.